

FIGURE 1

## FIGURE 2

HIV	RT	VKLPGMDCPVKVQ	WPLTEEKIKALVEICTEMEKECKISKICPKENPYNTPVFAIKKKDSTKWR
HTLV1	RT	RPWARTPPKAPRNQ	PVPFKPERLQLAQHLVRKALEAGHIEPYTG PGNNPVPVKA NGTWR
Ec-67	RT	HWLYRICSDNQTYTQFTI	PKGKCVTRISATPTRL KDIQRRCIDLLS DCRDEFAIRKI SNNYS
Mx-162	RT	AFHREVDATTHYSWTI	PKRDKCSKRTISPKPEL KAAQR WVLs NNV ERLP VHCAA
		o oooo ooo x x o o o	x o
HIV	RT	KLVDFRELNKRTQDFWEVQLGIPHAGLKK	KSVTVLVDVGDAYFSVPLDEDFRKYTA
HTLV1	RT	FIHDLRATNSLTDLSSSSPGPDLSSLPPTTLAHLQTIDLRAFFQIPLPKQFQPYF	A
Ec-67	RT	FGFE RGKSIILNAYKHRGKQII	LNIDLKDFFESFNFGRVRG YFLS NQDF
Mx-162	RT	HGFV AGRSILTNALAHQGADVVVKVLDKDFFSVTVRRVKGLLRRKGGLRECTSTLLSLLSTEAP	L
		oo o oo oo x o o oooooo x xx oo o x x	
HIV	RT	FTIP SINNETPGIRYQYQNVLPQCGWCKSFAIFQS	SMTKILEPFKKQNPDIVIYQYHDDLYVG
HTLV1	RT	FTVP QQCNYPGCTRYAWKVLQFCFKNSPTLFEM	QLAHILQPIRQAFPOCTIQLYHDDILL
Ec-67	RT	LN PUVATTLAAACYN GTLPGQSPCSPSIISNLICNIMDMLRALKKY	GCTYSRYYADDITI
Mx-162	RT	REAVQFRCKLLHVAKGP RALPQCAPTSPCITNALCLKLDRKLRSALAKRL	GFTYTRYADDITF
		o o ooooo o oo o o x x oo ooo x o x ooooo x	
HIV	RT	S DLEICQHRTKIEELRQHLLRWGLTP	DKKHQKEP PFLWMCYELHPDKWTVQPIVLPE
HTLV1	RT	S PSHEDLLLSEATMASLISHGLPVS	ENKTQQTIPGTIKEFLQGIIISPNHLTYDAPVTPVI RS
Ec-67	RT	STKNKTFPLEMATVQPEGVVLCKVLVKEI	ENSGFINDSKTRLYTQKTRQEVET CLTVNRVINID
Mx-162	RT	SWTAKQPKPRTQRQPPWAVLRSVQEVEAEFCRVHPDKTRVQRTRQRTV	GLVNUAGKDA
		o o o o o x o o ooo o oo o oo oo xx	
HIV	RT	SWTVDNIQKLVGKLNWASQIYP	
HTLV1	RT	RWALPELOALLGEIQWVSKGTP	
Ec-67	RT	RCYYKKTRALAHALYRTGE YK	
Mx-162	RT	PAARVPRDVRVQLRAAIHN RK	
		o	

FIGURE 3

REVERSE  
TRANSCRIPTASE

**A**

Mx-162	18	PTPELTAPSSDAAAKREARRLAHEALLVRAKAIDEAGGADDNVQAQLVSKGLAVEDLD-FSSASEKDKKA-WKEKK	91
Mo-MLV	1070	PDPDMTRVTNSPSLQAHLQALYLVQHEVW-RPL-AAAYQEQ-LDRPVVPHPYRVDTVVRRHQTKNLEPRWKGPY	1142
	o o o	o o o	o o o

Mx-162	92	KAEATERRALKRQAHEAW-KATHVGHLGAGVHWAEDRL	128
Mo-MLV	1143	TVLLTTPTALKVDGIAAWIHAHVKAADPGGG-PSSRL	1179
	o o o	o o o	o o o

**B**

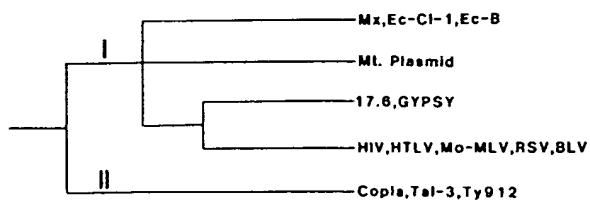
Mx-162	411	GKDAPAAPRVPDVVRQLRAAIHNRKKGKPGREGESLEQLKGMAAFIHMID-PAKGRAF-LAQLTELESTASAAPQAE	485
HIV	396	GKEGHHSARQCR-APR--RQGC--WKCGKPGHIMTNCPD-R-QAGFLGLGPWGKKPRNFPVAQVPQ-GLTPTAPP	461
	oo	oo o o o	o o o o o

Figure 4. Sequence Similarity of the msDNA-Mx162 Reverse Transcriptase with Other Retroelements  
(A) Sequence similarity of the region from residues 18 to 128 of the msDNA Mx162 RT (see Figure 2) with a carboxy-terminal region of integration protein of Moloney murine leukemia virus (M-MuLV) (residues 1070 to 1179; Shinnick et al., 1981)  
(B) Comparison of the sequence from residues 411 to 485 of the msDNA-Mx162 RT (see Figure 2) with the sequence from residues 396 to 461 of the gag protein of human immunodeficiency virus (HIV; Ralner et al., 1985).

FIGURE 4

**A**

Mx-162	304 GP-RALPQGAPTPGITA NACLKLDKRISALAKRL-GFTYTRVADELTF-SWTKA KQPKPRTQRPPVAVL	371
Ec-67	159 YN-GTLPOGSPCSP PIISNLICNIMDRILAKLAKKY-GCTYSRVAD ELITI-STNKN TFPLEMATVQPECVVL	226
Ec-86	130 YK-NLLPOGAPSSPKL ANLICSKLDYRIGGYAGSR-GLIYTRVAD ELITL-SAQSMKKVV KARDLFSIIPS	197
HIV	311 YQYNVLPQGKGSPA IFQGS---SMTKILEPFKKQNP DIVIYQYMDILYVG S-DLEIGQHRTKIEELRQHLL	377
HTLV1	150 YAWKVLPQGFKNS PTLFEM---QLAHILQPIRQAFP PQCTIQLYMDIL LLAS--PSHED LLLSEATMASLI	215
Mo-MLV	303 LTWTRLPOGFKNS PTLFDE---ALHRDLAD FRIQHPDL LILQLVDD LLAA-TSELD CQCG-TRALL-QTL	367
RSV	141 FQWKVLPQG MTCSPTICQL---VVGQV LEPLRLKHP PSLCMLH YMDIL LLAA-SSH DGLEA ACEEV I-STL	205
BLV	122 FAWRVLPOG FINS PALFER---ALQEPLR QVSAAF SQSLLV YMDIL LYAS--PTE EEQR SQCYQ ALA-ARL	186
Mt. plasmid	288 IATNGV PQGAST SGLATYN VL-----ELFLRY --DELI MAYADE GIL-CRQDP STPDFS VVEEAG VVQEP	348
17.6	339 YEYL RMPGLK NAP-ATF QRCMN-DI ---LRP LLNKHC-LV YLD IIIVFS-T SLDEH LQSLGL VFE--KL	399
GYPSY	284 YEF CRLP GFLRN ASSIF OR---AL DDV--L REQI-G KICV VV VDD VIIF S--EN ESDH VVR HID TVLK-CL	344
Copia	1032 CKLN KAIYGL KQAA RCW F- CIYI--LD KG NIN EN IYV-L LV VDD VIAT-- GDM TR MN NF KRY LME-KF	1112
Tal-3	990 CLL KKSL YGL LQSP RQW A-C VYV -K QVSE -Q EH LYL-- LL V DD MLIAG-- KSK SE INK V KE QL SM-EF	1069
Ty912	948 IRL KKSL YEL LQ -G AN W Y-E V R G-W SC V F K RS QV -T 1 CL F V DD V L F S-- K N L N S K R I I E K L K M-QY	1023

**B****FIGURE 5**

S 1 2 3 4



DEBETO 31 - DEBETO 32

FIGURE 6

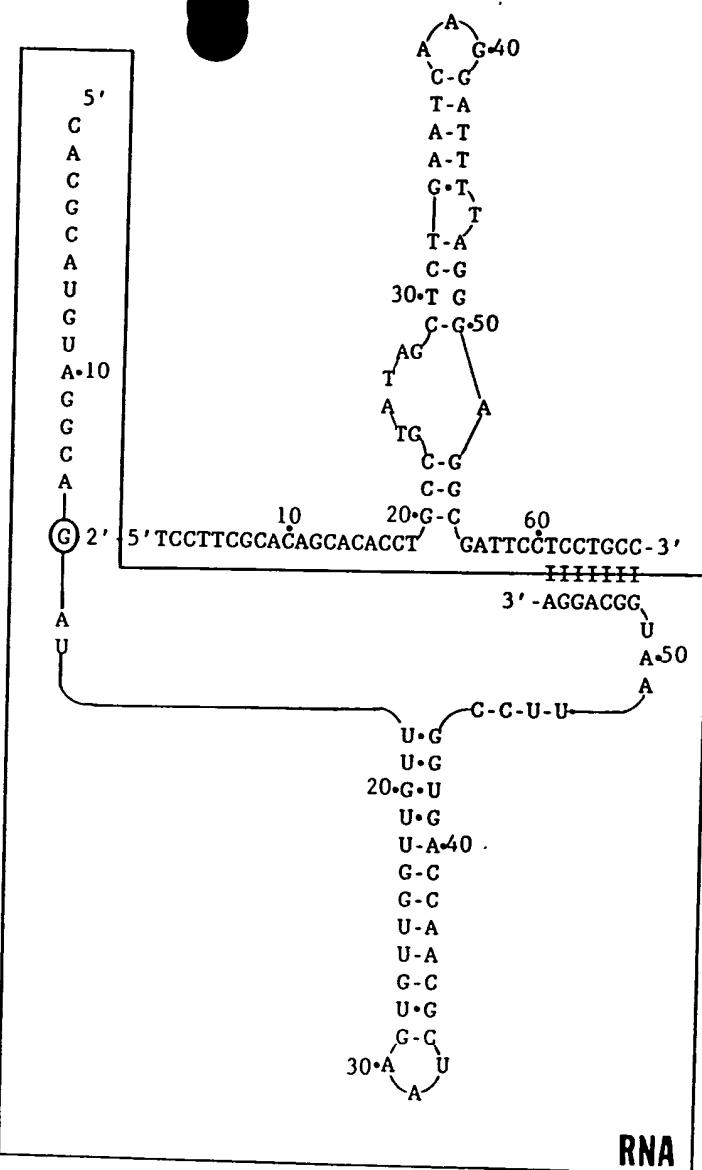


FIGURE 7

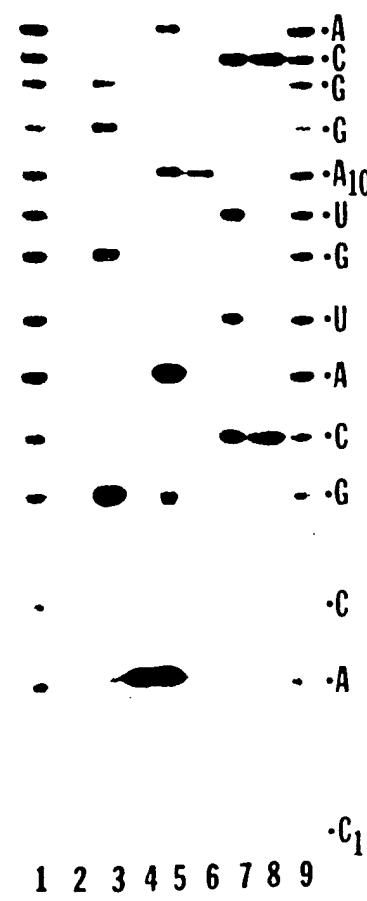
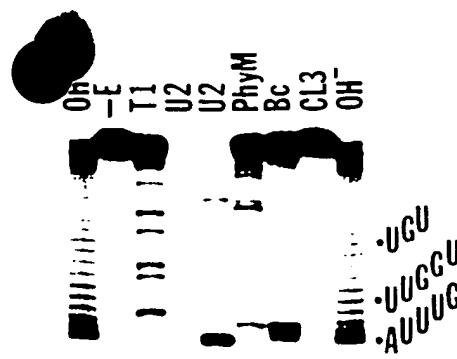


FIGURE 8

**A**

1 2 3 4 5

23.0-

9.4-

6.6-

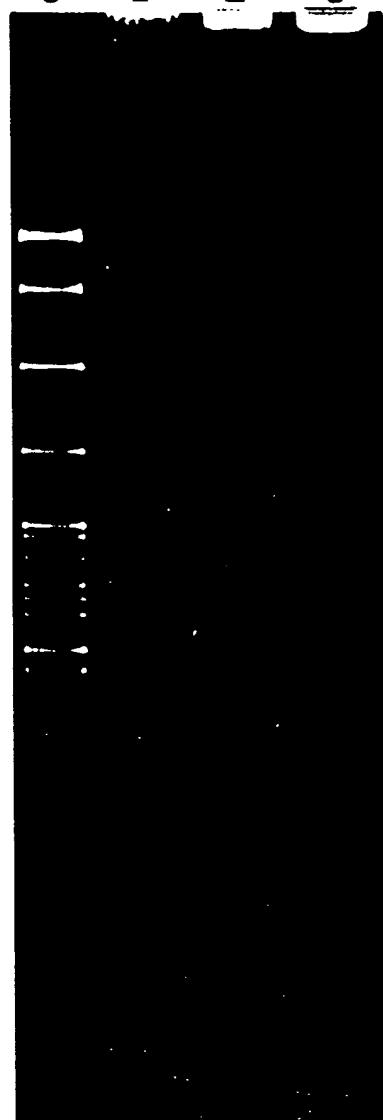
4.4-

2.3-

2.0-

**B**

S 1 2 3

**FIGU****FIGURE 9**

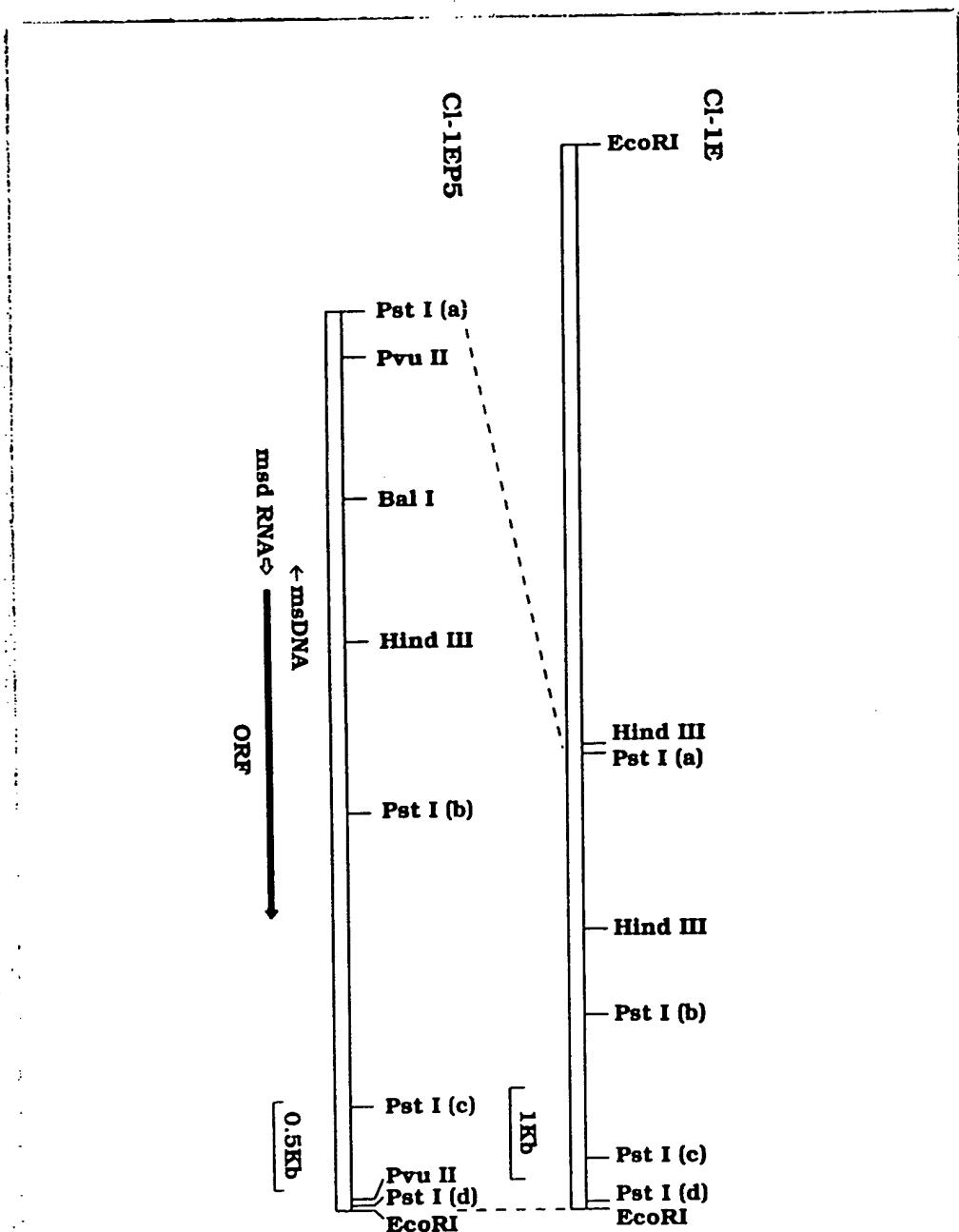


FIGURE 10

### FIGURE 11

V	RT	VKLKPGMDGPKVKQ	WPLTEEKIKALVEICTEMEMEKEGKISKIGPENPYNTPVFAIKKKDSTKWR	239									
LV1	RT	RPWARTPPKAPRNQ	PVPFKPERLQALQHLVRKALEAGHIEPYTG	PGNNPVFPVKKA	NGTWR	75							
DNA	RT	NVLYRIGSDNQYTQFTI	P KKGKVRTISAPTDRL	KDIQRRICDLSDCRDEIFAIRKI	SNNYS	94							
		+ o	● ●	+	● +	+							
V	RT	KLVDFRELNKRTQDFWEVQLGIPHAGLKKK	KSVTVLDVGDAYFSVPLDEDFRKYTAFTIP	SI	302								
LV1	RT	FIHDLRATNSLTIDLSSSPGPPDLSLPTTLAHLQ	TIDLRDAFFQIPLPKQFQPYFAFTVP	QQ	139								
DNA	RT	FGFE RGKSIILNAYKHRGKQIILNIDLKDFFESFNFGVRG	YFLS NQDF LLN	PVVA	150								
		o	●	+	+	o							
V	RT	NNETPGIRYQYNVLPQGWKGSPAIFQS	SMTKILEPFKKQNPDIVIYQYMDD	LYVGS	DLEIG	363							
LV1	RT	CNYGPGTRYAWKVLQPQGFKNSPTLFEM	QLAHILQPIRQAFPQCTILOQYMDDILLAS	PSHE	199								
DNA	RT	TTLAKAACYN GTLPQGSPCSPIISNLICNIMDMRLAKLAKKY	GCTYSR	<del>YADDITI</del>	STNKNTF	212							
		●	●●●	●●	+	o							
V	RT	QHRTKIEELRQHLLRWGLTTP	DKKHQKEP PFLWMGYELHPDKWTVQPIVLP	KDSWTVNDI	424								
LV1	RT	DLLLSEATMASLISHGLPVS	ENKTQQTPGTIKFLGQIISPQHLYDAVPTVPI	RSRWALPEL	262								
DNA	RT	PLEMATVQPEGVVLGKVLVKEIENSGFEINDSKTRLTYKTSRQEVT	GLTVNRIVNIDRCYYKKT	276									
		o	●	oo	o +	●	+	to					
V	RT	QKLVGKLNWASQIYPGIK	VRQLCKLLRGTKALTEVIPLT	EEAELELAENREILKEPVHGYYD	487								
LV1	RT	QALLGEIQWVSKGPTLRLQPLHS	LYCALQRHTDPRDQIYLNPSQVQSLVQLRQALSQNCRSRLVQ	327									
DNA	RT	RALAHALYRTGE YKVPDE	NGV LVSGGLDKLEGMF	GFIDQVDFKNNIKKKLNQ	PDRYVL	335							
		oo	+	+	o +	oo	●	o+o					
V	RT	PSKDLIA EIQQGQGQWTYQIYQE	PFKNLKTGKYARMRGAHTNDVKQLTEAVQKITT	544									
LV1	RT	TLPLLGAIMLTLGTTTVFQSKEQWPLVWLHAPLHTSQCPWGQLLASAVLLLKYTLQSY	GL	391									
DNA	RT	TNATLHGFKLKL NAREKAY SKFIY YKFFHGNTCP	TIITEGKTDRIYLKAALHSLET	SYPEL	396								
		o	●	oo	+	o	o	+	to	+	oo	o	
V	RT	ESIVIWGKTPFKLPIQKETWETWWTEYWQATWI	PE WEFV NTPPL	VKLWYQ	595								
LV1	RT	LCQTIHHNISTQTFNQFIQTS	SDHPSVPILLHHSHRF	KNLGAQTGELWNTFLKTAAPLAPVKALMP	456								
DNA	RT	FREKTDSKKKEINLNIFKSNEKTKYFLDLGGTADLKKF	VERYKNNYASYYGSV	PKQPVIMVLD	460								
		+	+	oo	+	o	o	●	oo				
V	RT	LE KEPIV GAETFYV	DGAANRET	KLGAGYVTNKGQK	VV PLTNTTNQ	KTELQAIYLA	652						
LV1	RT	VFTLSP VIINTAPCLFSDGSTSRAAYILWDKQILSQRS	FP LPPP	HKSA Q	RAELLG	LLLHGL	516						
DNA	RT	NDTG PSDLLN FLRNKVKSCPDDVTEM	TKM	KYIHVFY	NLYIVLTP	PLSPSGEQTS	MEDLF	PKDIL	523				
		o	●	o	+	+	+o	+	o +	●	o	●	o
V	RT	LQDS GLE VNIVTDSQYAL	QIIQA	QPDKSESELVNQII	IEQLIKKEKVYLA	WVPAHKG	708						
LV1	RT	SSAR SWR CLNIFLDSKYLYH	LRTLALGT	FQGRSSQAPFQA	LLPRLLSRKV	VVLYH	VRSHTN	578					
DNA	RT	DIKIDGKKFNKNNDGDSKTEY	GKHI	FSMR	VV RDKKRKIDF	KAFC	CIFDA	572					
		+	●	●●○	○ +	o	too	+					
V	RT	IGGNEQVDKLV	SAG				722						
LV1	RT	LPDPISRLNALT	DA				592						
DNA	RT	IKDIKEHY	KLMLNS				586						
		+	o	++									

FIGURE 12

**B**      **A**

**M-MuLV**

**pGB2**

**pCI-1EP5**

**FIGURE 13**

FIGURE 14

FIGURE 15

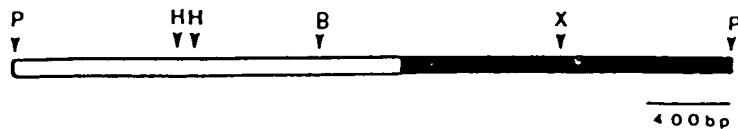
CC ACT TCC GGC GCT CGG GCT GCG CCA CGG CCC GTG CGA GCA CAT GAT GGC GCT GCG GCT 60  
 GT CCA GGT CGG GCA CGG CCC CCA GCA GGA ACC ACT CGG TCA GAC CCC CGG CGG CGA 120  
 CT CAT CGG CGG GCA GAC CGG CTC CTA CGT CGG CGG CGA GCA CGG CTC CGG CCA GCA GGT 180  
 TA CGG CGT CTC ATT CGA TGG GAA ACT GGT CGG GGT CGA CGT CGG CGG CGC CGA CGG CGA 240  
 TC CGG CGG GCA GAA GGT CGT CGT CGA CGC CGA CGG CGC CGC CGA CGC CGA CGA CGA 300  
 CG CGT CGG CGA CGG CGC CGG CGG ATA TAT CGA TGG CGC TGC CGT AAC GAT GAT GAA 360  
 AC GCA AGC CGC CGG CGC CGG CGG CGG CGG CGG CGG CGG CGG CGC CGA CGC CGA CGC 420  
 RNA<sup>32</sup> 480  
 AT CGT CGG CGC GAC CGA CGA CGA CGG AGG TCC CGA CGC ATC AGG CGC TCC ACC CGG CGG CGA 540  
 TA CGA CGC CGG CGG CGC CGG 600  
 CG CGG TTG CGC CGG CGC TCT CGT CGA ATT CGA CGG CGC CGG CGC CGG CGC CGG CGC CGG 660  
 CG CGC AAC CGG CGG CGA 720  
 T CTC TTG CCT CGG CGC TCC CGT CGG CGG CGG CGC CGG CGC CGG CGC CGG CGC CGG CGC 780  
 CG CGC CGG CGC 840  
 CG CGC CGG CGC 900  
 CG CGG CGG CGC 960  
 CG GAA CGG CGG CGC CGG 1020  
 CG CGG CGA CGG CGT GAC TCC CGG CGC CGG CGC CGG CGC CGG CGC CGG CGC CGG CGC CGG 1080  
 AG AGG AGG CGC CGG CGC 1140  
 AG CGC 1200  
 AC AGG ATC CGC CGG CGG CGG CGG CGC 1260  
 TG GAC TCC GTC GAG CGG CTG CGC AAA CGG CTG CGC ATC TCC CGT CGG CGC CGC CGC CGC 1320  
 CG CGC CGG CGC CGG CGC 1380  
 CG CGC CGG CGC CGG CGC 1440  
 1500  
 CGG CGG CGG CGC TCC ATC CTC ACC AAC CGG CTG CGC CAC CGC CGG CGG CGC CGC CGC CGC 1560  
 CGT GAC ATG AAG GAC TTC CGT CCT CGT ACC TGG CGC CGG GTC AAG CGA CGT CGT CGT CGC 1620  
 CGG CGA CGA CGA CGC CGG CGC 1680  
 CGG CGG CGG CGG CGC CGG CGC 1740  
 CGG CGG CGG CGC 1800  
 CGG CGT CGT CGG CGG CGC CGG CGC 1860  
 CGG CGC CGG CGC 1920  
 CGG CGT CGG CGC CGG CGC 1980  
 CGC CGG CGC CGG CGC 2040  
 CGT AAC GAG CGC CGC CGC CGG CGT CGG CGT CGC CGG CGC CGC CGC CGC CGC CGC CGC CGC 2100  
 CGT CGC CGG CGG ATC CGC AAC CGG CGA CGG CGC CGG CGC CGC CGC CGC CGC CGC CGC CGC 2160  
 CGG CGA CGG CGC CGC CGC 2220  
 CGG CGG CGC CGC 2280  
 CGG ATT CTG CGC CGG CGT GAA GAA GAC TTC CGA CGC CGG CGC CGC CGC CGC CGC CGC CGC 2340  
 TGA CTC CTC CGC CGG CGC 2400  
 CGC 2460  
 CGC 2520  
 CGC 2580  
 CGC 2640  
 CGC 2700  
 CGC 2760  
 CGC 2820  
 CGC 2880  
 CGC 2940  
 CGC 3000  
 CGC 3060

FIGURE 16

TTTGAGAAG CGCCATACCA AACAGGGAT ACAGACCAAC CTGACGCTGA AAGAGGAAG CTAGGCCCAC TGGCTCCCGA AGTCGCCAGA 9930  
 T E K R H T R Q G I O T H L T L K E S T G D V L P K C D  
 CCCGCCACCA ACATAACCTC ACTCAAGCCG GAAACAGCCG GTCTTTCCTC TTCTGCGCAT TCCCACAGG TGAAAGATCC ACTTCAGCA 10080  
 P A A T  
 CTGACCCCT TATTCACATG CAAATTATTA ATAAGAAACG AGAGCTGACG AGTCGCCAGA GAAAACCTG AAAAAGCTT 10170  
 TTTATCACCCG CGGCATGCCG CGACTGGACA GATCCAGAAC GACCAAAAT CACAAAGCTG AGCAGTCAC TGTTCAGCTC TCAACCAACTC 10260  
 ATCACCACCT AACCACATG AATAAAATG AGCTTACGCG TAAATGCG AAAAAGCTG TTCTGCGCAT TGGATAAG 10350  
 AAAAATTAAT CACATCAATA GCTTTCCTC TGAACTCTC TGAGGTTTATG GAGGCGTAA ~~AGAGCCAA~~ CTTTCAGTT TATGGTTA 10440  
 TTTAATTAAG GTGAGTTAT CAAAGGAGA ACTTAGGAGA ACTTCAAAATA CTCTCCATTG TTCTGCGCAT TTCTGCGTTT GGATGTTAA ATACCCATT  
 TGGCCATCC CGCAGTCAGTCGCTC CTAGCTTCCGCTC TTCTGCGCAT TGGCTTACGCTC TGACTTCAGCTC 10530  
 ATCGGCTTCCG CGCTACTCG TACCAAGG GATCATTTAA ~~GGCTACCGG~~ CAGGAGCTCG AGGACTCGC CGCCGACCC CGCTGCGCTAG  
 AGCGAACTGA TGAGCTCTC CTAGCTGGT TTGGCTCTT AGTCTCTAC CTCACAGCTG CATAAGGATA TTCTGCGCAT TGACTTCAGCTC 10620  
 TCCCTTGACT AGCTGAGCCG GTTCATCCAA ACCGAGAAA TGAGGAGATG GTAGTTCTAC GATTCCTAT AGAGCTTACG ACTGAGCTCG  
 Of 314 N L T Q L  
 > DNA  
 AAAAAAAAT GGTACTGAGG TATTCACCCG AACCCGTTA TTTCATCATG TGTTGAAAGA GAAACAGTA AAATGCTCTG GAAATGTTAAA 10710  
 K E K R G T E V B R A T A L P E S F V V E K K C P G H V K  
 AAAATTCCTC TTTCCTGCTG GTGCTAACTA AAAACATGCA GAAACATCAG CAACAGCATTG GCAAAATAA AATTTTCCTG AAAGGTATTG 10800  
 K F V P L C G A N K H H G E P S A R R L E L I H F S E K V L  
 GAAATACTG CACTTTTG TTGGCTGAACT AGTTTTCCTA GAAATTAAGCA CGGATGAGA ATCATTATCT GATAATGCA 10890  
 H N C H F Y L A E L V P K E L S T D E E S L S D H L L D I E  
 AGCTGACTTA TCTAAATTAC CTGATCATAT TATCATTGTT TTAGAAGTT ATTACATCTT CACGGACTG CGTCATTCG CATAAGCGAA 10980  
 A D L S K E A D H I I I V L E S Y S E P T E L G A F A Y E K  
 GCAATTACCG AAGAAATTAA TAATAGTTAA CAATACAAA TTTCATATACT AGAAATCTC TATAAATATG GCAAAATAA AGGCTTATTAC 11070  
 Q L R K K E L I V I N H M T E F I N E K S Y I N H G P I K A I T  
 TCAGCAATCA CAACAACTG CTGCTTCTC ACATTATAAA ATGACAGAGG GTATTGAAAG TATAGGAGG TGTCATGGG TTGGCGAAAT 11160  
 Q Q S Q Q S G H F L H Y E N T E G I E S I E R S D G I G E I  
 ATTCGACCC CGTATGATA TTCTCTCA GAAACAGACA GCAATTCTCA GAACTTTAA AAAACAGAG TGAGCTCTT CGCTAACTT 11250  
 F P O L Y D I L E K H D R A I S E R T L K E E L D P E S S V Y  
 CAATAAAGAC TCAGTACGAT TTTCATGAGC GCAATTCTCA GAACTTTAA AAAACAGAG TGAGCTCTT CGCTAACTT 11340  
 N K D S V R F I N D O V I F V C G P L Q L H E L I K I I T K I  
 ATTTGGCACA GAAAGCCATT ACAAAAAAAAT TTCTCTAAAG CACTCTGTTA TTCTAATAGC TATTAGATAA ATATCATGCA CAATGGGAT 11430  
 F G T E S H Y K E N L L E K H G I L I A R I I E C T N G I  
 TTATATCTT TTGTATAAG ATATTTATTTA TAATATGAC TTTCAGATG ACACATCTC ATCAATGTTT AAAGTTTTT TTCTCAAGAA 11520  
 Y Y S L Y Y Y P V Y D C P D I D H M I S H N F K V F E L K N  
 CAAGCCAGAA AGGATGAGGG TATATGAGAA TATATAGCTT ATTTGATTCT CACACATGCA TGACTAAGGG ATTTGCTCTC GAAGTAATGC 11610  
 K P E R H R K V Y E N I \*  
 RT H R I Y S L I D S Q T L H T K G F A S E V N  
 GATCAGCTGA CGCGCCAAAAT AACGGGATA TAGCTAAGAA AAAGGGGT ATGACAAACAA TTTCATGCCG GTCATCAAATGTTAAATTTAA 11700  
 R S P E R K K M D I A K K K G G H R T V E E K V K L  
 TTCAATATG ATATTTTTT CGAAGCTCCG ATTCATCAATG GTCTGCTATG CATTTCCTA AAACGGATCA ATAAAAGCA 11790  
 I Q Y W L K H H H V T S K L P R H H A A Y A F V K H R S I K S  
 ATGCTTTTACATGCCAA TCAAGAATA ATGTTATGCT GAAATAGATG CTCAAGATG TTTCCTCTC ATAAAATTT ATGATTTTG 11880  
 H A L L H A E S K H V V K I O L K D O F F P S I E F T D F  
 AGTAGCCAT CACTCTTAT CGAGATGCA TTGAATTCG TACACATAT GATAAGGACT TACTACACT TATAAAAGG ATTCGCTTTA 11970  
 E Y A F T T R D R I E F T T E Y D K E L L O L I K T I C F  
 TATCAGATG CACTCTCCCT ATCGGGTTTC CTACATCTCC ATTAAATGCA ACCTTTGGG CAAGACACT TGATGAAAAA CTGACGCAA 12060  
 I S D S T L P I G F P T S P L I A H F V A R E L D E K L T Q  
 AACTAAATGCA ATTTGATAAA CTAAATGCCA CTTATACAGG ATATCTCTGAT GATATTTATG TTCTCTACAA TATGAAAGG GCTAGCAAAT 12150  
 K L N A I D K L H A T A D D D I V S T R H E G A S E  
 TAATCTGGA TTGTCTTAA AGAACATGA AAGAGATGG TTCAAGCTT AAAATTACAA TTAAAATTA TAATGATTGT ATGCTCTGG 12240  
 L I L D C F E R T H K I G P D F K Y E I E K F K I C S A S  
 GAGGAAGTAT AGTAGTTACG GGATGAAAG TTGGCCACCA TTTCATATT ACATTCATCA GATCAATGAA AGATAAAATA AGATGCTAC 12330  
 G G S I V V T G L K V C H O F H I T L H R S K R D K I R L H  
 TTCTCTTTT ATCAAAGGGC ATATTTAAAG ATGAGATGCA TAATAAACTT TTGGTTATA TTGGCTTATG AAAGATATA GACCCCTTATT 12420  
 L S U L L B K O I L K D E O H N K L S G T I A Y A K D I D P H  
 TTATACAAA ATGACACAGA AAATTTTCG AAGAAATATA ATGAGATTCG ATATCTCACCA ACAGAGTGA ATAAACTTAA TATTTGGAT 12510  
 F Y T K E L H R K Y F Q E I K W I Q H L H H K V E \*  
 GCACCCCAAT AACCTCATG ATTAATTTGG GAAACATATAA GCGCTTTCAG GATGACTAC ACTCTAGAGA ATGCTTACG AAAAGCTTAT 12600  
 AAGTTATTTT CAACCTATAA TAAATACAG CAAACATCAAT GCAATGGGGG CATTTCACG ~~CTTCCTGCTGAT~~ CTTCCTGCTAATGCTCA 12688

FIGURE 17

(A)



-371 TGGCATCTATTAAGAAGGTTAGGAAAGAAAAATAAAGTATCAAAGATATTGAAATATAT  
 -311 TATACGCAGAGCGTTCTATGCCCTGTATCTATTACTGGATAGTGTCAACTACCGCAC  
 -251 ACTGTGTGAACTAGCTTTAAGCGATAAAGCAAGATGATGTTTATCTAAAATTATTGT  
 -191 TAGATCCGTTGTTCTCGCTAATAAAATGAAACGAAAAATACTCAATGACTGATGGTTA  
 -131 TCAGGTCACTGCTTGGGGCTAGCTATGTTAGGAGGCTTGTAGAGAAAGACACTTGA  
 -71 CCGATTGCGGCTTGAGATTATGAATTGAAACCGTAGAAATCAACATTTAACTATGA  
 +1  
 -11 TAAGATTCCGTATGCGCACCCCTAGCGAGAGGTTATCATTAAGGTCAACCTCTGGATGT  
 IR ----->  
 49 TGTTCGGCATCTGCATTGAATCTGAGTTACTGCTCTGTTCTGTTGAAACGGAGAG  
 <-----  
 109 CATGCCCTGATGCTCTCCGAGCCAAACCAAGGAAACCCGTTTTCTGACGTAAAGGTGCGC  
 bDNA-----|-----  
 169 AACTTCATGAAATCCGCTGAATATTGAAACACTTTAGATTGAGAAAATCTCGGCCCTACC  
 - IR MetLysSerAlaGluTyrLeuAsnThrPheArgLeuArgAsnLeuGlyLeuPr  
 229 TGTCTATGAAACAATTTCATGACATGTCTAAGGCAGCTGCATATCTGTTGAAAACACTCG  
 ovalMetAsnAsnLeuHisAspMetSerLysAlaThrArgIleSerValGluThrLeuAl  
 289 GTTGTAACTACAGCTGATTTCGCTATAGGATCTACACTGTAGAAAAAGAAAGGCC  
 gLeuLeuIleTyrThrAlaAspPheArgTyrArgIleTyrThrValGluLysGlyPr  
 349 AGAGAAGAGAATGAGACCATTTACCAACCTCTCGAGAACTTAAGCCTACACGGATG  
 oGluLysArgMetArgThrIleTyrGlnProSerArgGluLeuLysAlaLeuGlyTr  
 409 GGTCTACGTAACATTAGATAACTGTCGTACATCCTTTCTATTGATTTGAAAA  
 pValLeuArgAsnIleLeuAspLysLeuSerSerProPheSerIleGlyPheGluLy  
 469 GCACCAATCTATTGAAATACTGCTACCCCGCATATTGGGCAAACTTATACTGAAATAT  
 sHisGlnSerIleLeuAsnAsnAlaThrProHisIleGlyAlaAsnPhelleLeuAsnIle  
 529 TGATTGGAGGATTTCACAGTTAACGCTAACAAAAGTTTGGAGTGTTCCTTCATT  
 eAspLeuGluAspPhePheProSerLeuThrAlaAsnLysValPheGlyValPheHisSe  
 589 TCTGGTTATAATCGACTAAATCTTCAGTTTGACAAAAATATGTTGTTATAAAAATCT  
 rLeuGlyTyrAsnArgLeuIleSerSerValLeuThrLysIleCysTyrLysAsnLe  
 649 GCTACCACAAAGTGTCTCCATCATCACCTAAATTAGCTAACATCTAAATGTTCTAAACTTGA  
 uLeuProGlnGlyAlaProSerSerProLysLeuAlaAsnLeuIleCysSerLysLeuAs  
 709 TTATCGTATTCAAGGTTATCCAGGTAGTCGGGCTTGTATATACGAGATATGCCGATGA  
 pTyrArgIleGlnGlyTyrAlaGlySerArgGlyLeuIleTyrThrArgTyrAlaAspAs  
 769 TCTCACCTTATCTGACAGCTATGAAAAAGGTTGTTAAAGCAGTGTGATTTTTATTTTC  
 pLeuThrLeuSerAlaGlnSerMetLysLysValValLysAlaArgAspPheLeuPheSe  
 829 TATAATCCCAAGTGAAGGATTGGTTATTAACCTCAAAAAAAACTTGTATTAGTGGGCTCG  
 rIleIleProSerGlyLeuValIleAsnSerLysThrCysIleSerGlyProAl  
 889 TAGTCAGAGGAAAGTTACAGGTTAGTTATTTCACAAAGAGAAAAGTGGGATACGTTAGAGA  
 gSerGlnArgLysValThrGlyLeuValIleSerGlnGluLysValGlyIleGlyArgGly  
 949 AAAATATAAGAAATTAGCACAAGATACATCATATAATTGCGGTAAGTCTCTGAGAT  
 uLysTyrLysGluIleArgAlaLysIleHisHisIlePheCysGlyLysSerSerGluIle  
 1009 AGAACACGTTAGGGGATGGTTGTCATTATAATTAACTGTGTTGATTCAAAAGCCATAGGAG  
 eGluHisValArgGlyTrpLeuSerPheIleLeuSerValAspSerLysSerHisArgAs  
 1069 ATTAATAACTTATATTAGCAAATTAGAAAAAAATATGAAAGAACCCCTTAAATAAAAGC  
 gLeuIleThrTycIleSerLysLeuGluLysLysTyrGlyLysAsnProLeuAsnLysAl  
 1129 GAAGACCTAATGGCTTCGTTAAACATAAGCTCATAGGTTGAAAGAACCCACTTC  
 aLysThr  
 1189 TTCTGTCACCAACCAGTTATTAGTTCCTGCAATCGTTCTGCAG

FIGURE 18

Oligo 2337  
tcaccctgaaagac~~t~~gattgcttacctggaagagaagccgaaatggccgaacatctgg 60  
cgcggttaaggctatcgcaagagttcgccgtttaaaATATGGCTCTGCAGGGTT 120  
**RNA** a2  
TTGCTGTCGCA~~CGT~~GATGGCTTCAGA~~TATCGT~~TTAATCTGCTT~~CGCCAGCAGTC~~ 180  
AACGACACGCGTCGCACTACCGCAAGTTCTATGCACAATTAGACGAAAGCGGTCTGTCAC  
CGAATAGCGTTCCGGCTTTGTGCCCCAGGGTCGGCCAGTCGCTGACTTAACGCCAG 240  
CGTTATCGCAAAGGCCAAAACCGCCCTCCAGCCGCTAGCGACTGAATTGGGTC  
TAGTATGTCATACCCAAAGTCGCTTATTGACCTGAGTACGCTTCGGTACGTCCG 300  
ATCATACAGGTATATGGGTTTACGGCAAGTAACATGGACTATGCGAACGCCATGCAGCG  
CGACTGCGGAGTCATGTCATGCGGAAAGCCCTACCAAATTACATAACGGGACAAAC  
GCCTCTTTGGCCCGTGTGGAGAGTGGATGGTACCCGGACAACCCCTC 360  
M D A T R T T L L  
TGGCGCTCGATTGTTCCGGCTCGCCGGGCTGGAGCGCGATAAGAAATACAGCGACTGC 420  
A L D L F G S P G W S A D K E I Q R L H  
ATGCGCTCAGTAATCATGCCGACGCCATTACCGACGCATTATTCTTCTAAACGCCACG 480  
A L S N H A G R H Y R R I I L S K R H G  
GTGGTCAGCGGCTGGTGTAGCCCTGATTACTTGCTAAAACCGTACAGCGAACATTC 540  
G Q R L V L A P D Y L L K T V Q R N I L  
TTAAGAACGTCCTTCACAATTCCGCTTCCCTTTGCTACAGCCTACCGGACAGGTT 600  
K N V L S Q F P L S P F A T A Y R P G C  
GCCCAATCGTACGCAACGCCAGCCACACTGCAACAGCGCAGATCTGAAACTCGATA 660  
P I V S N A Q P H C Q Q P Q I L K L D I  
TCGAAAACTTTCGATAGCATTAGCTGGTACAGGTCTGGCGTGTGTTGCCAGGCC 720  
E N F F D S I S W L Q V W R V F R Q A Q  
AGTTGCCACGTAACTGGTAAACCATGCTGACCTGGATTGTTATAACGACGCGTTAC 780  
L P R N V V T M L T W I C C Y N D A L P  
CGCAGGGGGCACCAACTCGCCAGCCATTCCAATCTGTGATGCCGTTTGATGAAC 840  
Q G A P T S P A I S N L V M R R F D E R  
GCATAGGGGAATGGTGTAGGCTCGGGGAAATTACCTACACCCGTA~~CTGGATGACATGA~~ 900  
I G E W C Q A R G I T Y T R Y C D D M T  
CCTTTTCAGGTCACTCAATGCCGCCAGGTTAAAATAAGTGTGCCATTGTTAGCGG 960  
F S G H F N A R Q V K N K V C G L L A E  
AGCTGGGCTGAGCCTCAATAACGCAAAGCTGCCTGATAGCTGCCGTGAAGCGCCAGC 1020  
L G L S L N K R K G C L I A A C K R Q Q  
AAGTAACCGGGATTGTTGTTAATCACAAGCCACAGCTGCCGTGAAGCGCCGGCGC 1080  
V T G I V V N H K P Q L A R E A R R A L  
TGCGTCAGGAGGTGCATTGCCCCAAAATGGCTTATTCGATCTTAGTCATCGTG 1140  
R Q E V H L C Q K Y G V I S H L S H R G  
GTGAACCTGATCCTCTGGCGATCTCCACGCAAGGCAACGGCGTATCTTATGCTTG 1200  
E L D P S G D L H A Q A T A Y L Y A L Q  
AGGGAAAGAATAAACTGGTATTGCAAATCAACCCGAGGGATGAGGCCCTTCAACAGCGA 1260  
G R I N W L L Q I N P E D E A F Q Q A R  
GAGAGAGTGTAAAGCGAATGCTGGTTGATGGTAAGAAAAGCGTCAGGCAGACGTTCTG 1320  
E S V K R M L V A W \*  
CCTGACCGTTAGGGAGAattactgcaactgcgcggcaattagcgccagcggcgtca 1380  
aaatcatccgtcgccgtat~~tt~~aaactcgctcgccgacaaaacgtgacagcatacctca  
cagaaggccaggatctggctgccagcagg~~ttt~~catcg 1440  
Oligo 2336 1500  
Oligo 2336 1540

FIGURE 19

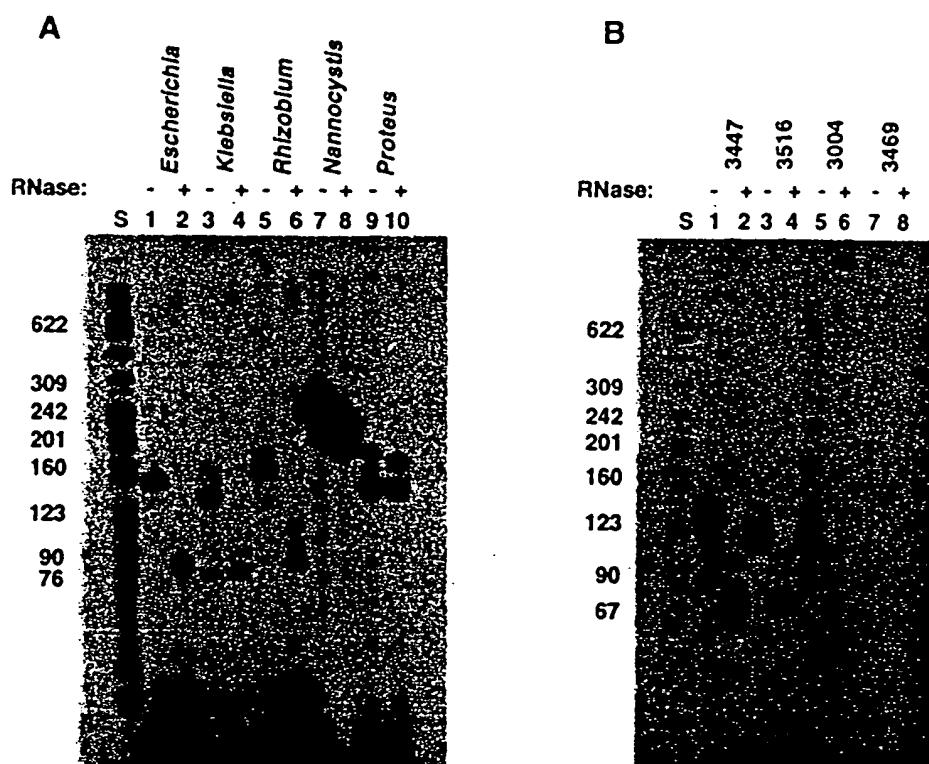


FIGURE 20

## RHIZOBIAL ISOLATES

Strain (legume host genus)	USDA strain no.	Geographic source (date)	msDNA produced <sup>b</sup>
<i>Rhizobium</i> sp. ( <i>Acacia</i> )	3002	Brazil (1959)	+
	3003	Africa (1950)	
	3325	Morocco (1974)	
	3838	? (1976)	
<i>Bradyrhizobium</i> sp. ( <i>Aeschynomene</i> )	3516	Florida (1972)	+
	4362		+
<i>Bradyrhizobium</i> sp. ( <i>Albizia</i> )	3004	Maryland (1952)	+
<i>Bradyrhizobium</i> sp. ( <i>Apis</i> )	3240	Maryland (1939)	
<i>Bradyrhizobium</i> sp. ( <i>Arachis</i> )	3339	Thailand (1979)	
	3341	Hawaii (1978)	
<i>Rhizobium</i> sp. ( <i>Astragalus</i> )	3854	Alaska (1962)	
<i>Rhizobium</i> sp. ( <i>Cajanus</i> )	3472		
<i>Bradyrhizobium</i> sp. ( <i>Canavalia</i> )	3317	Brazil (1974)	
<i>Rhizobium</i> sp. ( <i>Cicer</i> )	3378		
	3379	Mexico (1963)	
<i>Bradyrhizobium</i> sp. ( <i>Coronilla</i> )	3165	Virginia (1935)	
	3167	? (1961)	
<i>Bradyrhizobium</i> sp. ( <i>Crotalaria</i> )	3384	Brazil (1967)	
<i>Bradyrhizobium</i> sp. ( <i>Desmodium</i> )	3225	Ecuador (1948)	
<i>Bradyrhizobium</i> sp. ( <i>Erythrina</i> )	3241		
	3242	Maryland (1939)	+
<i>Rhizobium fredii</i>	191	China (1979)	
<i>Rhizobium leguminosarum</i>	2370	Illinois (1933)	
	2429	Hawaii (1978)	
	2435	Holland (1955)	
	2480	Tennessee (1951)	
	2489		
<i>Rhizobium</i> sp. ( <i>Lens</i> )	2426		
	3404	Colombia (1979)	
<i>Rhizobium loti</i>	3084	Maryland (1946)	+
	3468	New Zealand (1961)	
	3469		
	3471		
	3503		+
<i>Bradyrhizobium</i> sp. ( <i>Lotus</i> )	3669	California (1968)	
	3074	Minnesota (1954)	
	3470	California (1916)	
<i>Rhizobium</i> sp. ( <i>Lupinus</i> )	3040	Florida (1940)	
<i>Bradyrhizobium</i> sp. ( <i>Lupinus</i> )	3045	Florida (1946)	
<i>Bradyrhizobium</i> sp. ( <i>Macrotyloma</i> )	3451	Zimbabwe (1960)	
<i>Rhizobium medicago</i>	1097	North Dakota (1948)	
<i>Rhizobium meliloti</i>	1011	Maryland (1933)	
	1021a	North Dakota (1948)	
<i>Rhizobium phaseoli</i>	2667	Washington (1948)	
	2669		
	2674	Brazil (?)	
	2676	Colombia (1972)	
	3256	Illinois (1941)	
<i>Rhizobium</i> sp. ( <i>Robinia</i> )	3436		
<i>Bradyrhizobium</i> sp. ( <i>Stylosanthes</i> )	3441	Brazil (?)	
	3477	Colombia (1976)	
<i>Rhizobium trifolii</i>	2046	Virginia (1934)	+
	2048	Illinois (1934)	
	2063	Florida (1939)	
	2065	Alabama (1952)	
	2116	South Carolina (1944)	
	2134	? (1974)	
	2145		
<i>Rhizobium</i> sp. ( <i>Trigonella</i> )	2156	California (1920)	
<i>Rhizobium tropici</i>	1177	Florida (1939)	
<i>Bradyrhizobium</i> sp. ( <i>Vigna</i> )	2744	Brazil (?)	
	3447	Thailand (1979)	
	3456	Wisconsin (1966)	+

<sup>a</sup> All strains are from the USDA Beltsville Rhizobium Culture Collection, provided by Peter van Berkum.

<sup>b</sup> As defined by detection of radiolabeled msDNA by the RT extension method.